SPECIAL REPORT

OPTICAL SYSTEMS GROUP

DIGITAL MOTION IMAGERY COMPRESSION BEST PRACTICES GUIDE – A MOTION IMAGERY STANDARDS PROFILE (MISP) COMPLIANT ARCHITECTURE

WHITE SANDS MISSILE RANGE
REAGAN TEST SITE
YUMA PROVING GROUND
DUGWAY PROVING GROUND
ABERDEEN TEST CENTER

NAVAL AIR WARFARE CENTER WEAPONS DIVISION, PT. MUGU
NAVAL AIR WARFARE CENTER WEAPONS DIVISION, CHINA LAKE
NAVAL AIR WARFARE CENTER AIRCRAFT DIVISION, PATUXENT RIVER
NAVAL UNDERSEA WARFARE CENTER DIVISION, NEWPORT
PACIFIC MISSILE RANGE FACILITY
NAVAL UNDERSEA WARFARE CENTER DIVISION, KEYPORT

30TH SPACE WING
45TH SPACE WING
AIR FORCE FLIGHT TEST CENTER
AIR ARMAMENT CENTER
ARNOLD ENGINEERING DEVELOPMENT CENTER

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

DISTRIBUTION A: APPROVED FOR PUBLIC RELEASE
DISTRIBUTION IS UNLIMITED

This page intentionally left blank.

DIGITAL MOTION IMAGERY COMPRESSION BEST PRACTICES GUIDE – A MOTION IMAGERY STANDARDS PROFILE (MISP) COMPLIANT ARCHITECTURE

June 2012

Prepared by

OPTICAL SYSTEMS GROUP
(OSG)

Published by

Secretariat
Range Commanders Council
U.S. Army White Sands Missile Range
New Mexico 88002-5110

This page intentionally left blank.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

PREFACE

This document provides guidance for the DoD (Department of Defense) Test and Evaluation (T&E) Range community on best practices for how to implement and perform compression and streaming on range digital motion imagery. This includes all types of motion imagery applications at the ranges from low-latency situational awareness monitoring to compression of high-speed, high-resolution, high-bit depth, and infrared imagery for processing, customer delivery, and archival purposes. These practices are based on a Motion Imagery Standards Profile (MISP) compliant architecture, which has been defined and approved by the DoD Motion Imagery Standards Board (MISB). While the MISB focuses on warfighter motion imagery requirements, there are several reasons, justifications, and advantages for the T&E ranges to adopt this architecture. Interoperability between tactical and range systems and formats, leveraging of technology developments, availability of software tools and libraries, commonality of metadata formats, and the ability of programs to readily ingest range test data for analysis purposes are to name but a few. This MISP is a living document that is frequently updated by the MISB. The MISB website is www.gwg.nga.mil/misb. Motion imagery system implementers are encouraged to monitor changes to the MISP and to leverage new formats, profiles, and technologies as they are adopted.

For development of this report, the RCC gives special recognition to:

Task Lead: Dr. Joe Stufflebeam
White Sands Missile Range
TRAX International
PO Box 398
White Sands Missile Range, NM 88002
Phone: (575) 678-7917
Email: joseph.l.stufflebeam.ctr@mail.mil

Please address any questions to:

Secretariat, Range Commanders Council
ATTN: TEDT - WS - RCC
100 Headquarters Avenue
White Sands Missile Range, NM 88002
Phone: (575) 678-1107 DSN 258-1107
Fax: (575) 678-7519 DSN 258-7519
Email: usarmy.wsmr.atec.list.rcc@mail.mil

This page intentionally left blank.

ACRONYMS

AAF	Advanced Authoring Format
AVC	Advanced Video Codec (H.264)
B	Bi-directional
BP	Baseline Profile
CAVLC	Context-adaptive Variable-length Coding
CBP	Constrained Baseline Profile
DCT	Discrete Cosine Transform
DoD	Department of Defense
DWT	Discrete Wavelet Transform
EG	Engineering Guideline
FRExt	Fidelity Range Extension
GOP	Group of Pictures
HD	High-definition
HiP	High Profile
Hi10P	High 10 Profile
Hi422P	High 4:2:2 Profile
Hi444PP	High 4:4:4 Predictive Profile
IBBP	Interframe-Bidirectional-Bidirectional-Predictive
JPEG	Joint Photographic Experts Group
Mb	Megabits
Mbps	Megabits per second
MISB	Motion Imagery Standards Board
MISM	Motion Imagery System Matrix
MISP	Motion Imagery Standards Profile
MPEG	Moving Picture Experts Group
MTF	Modulation Transfer Function
MXF	Material Exchange Format
n.d.	No Date
OSG	Optical Systems Group
P	Predictive
PSNR	Peak Signal to Noise Ratio
RF	Radio Frequency
RGB	Red Green Blue
RP	Recommended Practice
SMPTE	Society for Motion Picture and Television Engineers
STD	Standard
T&E	Test and Evaluation
TS	MPEG-2 Transport Stream
TSPI	Time Space and Position Information
UTC	Universal Coordinated Time
XP	Extended Profile

This page intentionally left blank

CHAPTER 1

INTRODUCTION

There are two main components to the Best Practices recommendations that are made in this document. The first defines the desire to build range T&E motion imagery systems that are compliant with the standards-based DoD architecture. The goal here is centered on maximizing interoperability with motion imagery content. The second outlines a number of issues and mechanisms for configuring and tuning a system for optimal performance under certain given conditions.

As compliance with the DoD Motion Imagery architecture is a primary goal, it is first appropriate to define what is meant by compliance. The definition of MISP Compliance is provided in the MISP as follows.

> **MISP COMPLIANCE Definition:** Motion Imagery Standards Profile (MISP) compliance is based upon compliance to a specified approved version of the MISP (e.g. MISP Version (V) 4.4, MISP V4.5, etc.). The motion imagery system supplier specifies the MISP version for which it is seeking compliance along with three qualifications: the [Motion Imagery System Matrix] MISM-Level for video compression, the file format for transport or storage, and the metadata RPs [Recommended Practices] / EGs [Engineering Guidelines] / STDs [Standards] used. MISM levels are as defined per the MISP version specified by the system supplier. All signals tested are assumed digital. Supported video compression includes Moving Pictures Expert Group (MPEG)-2, MPEG-4 Part 10 (i.e. AVC [Advanced Video Codec] or H.264), and Joint Photographic Experts Group (JPEG) 2000. Supported file formats include MPEG-2 Transport Stream [TS], MXF [Material Exchange Format], and AAF [Advanced Authoring Format]. Furthermore, if the motion imagery system uses MXF/AAF it shall comply with Standard 0301. Metadata is tested for compliance to the specified version of the MISP and respective EGs/RPs. Draft RPs/EGs will not be tested until approved by the MISB. MISP compliant systems shall produce metadata elements from Standard 0601 or EG 0104 (legacy systems only), optionally using metadata keys from MISB Standard 0807, SMPTE [Society for Motion Picture and Television Engineers] RP 210, and MISB RPs/EGs of their choice. In addition, Security metadata shall comply with MISB Standard 0102.[1]

The key takeaway of the MISP Compliance definition for this document is that currently-approved video compression formats include MPEG-2, MPEG-4 Part 10 (H.264), and JPEG 2000. As MPEG-2 is supported as a legacy format only, H.264 and JPEG 2000 are therefore the video compression formats that should be leveraged for new range T&E compression applications. The choice of whether to use H.264 or JPEG 2000 in a particular application is driven mainly by the application and the format of the source material. In general, H.264 tends to be applied to motion imagery streams that are based on a broadcast video format, such as

---

[1] MISB (Motion Imagery Standards Board). Motion Imagery Standards Profile v6.2. Accessed June 2011. Available at http://www.gwg.nga.mil/misb/misppubs.html.

standard-definition or high-definition video. The JPEG 2000 format tends to be applied to applications where the imagery is of higher resolution (for instance > 2 Mb per frame), such as digital cinema, film scanning, etc. There is not a hard-line decision on when to use H.264 versus JPEG 2000. The compression efficiency of the two algorithms for intraframe compression is similar. The manifestation of compression artifacts is different and preference tends to vary among end users. Examples of how the two algorithms tend to break down are provided later in the document.

CHAPTER 2

IMAGE FORMATION

Image compression is a process by which the amount of data used to represent an image or sequence of images is reduced. The purpose for reducing the size is to gain efficiency when transmitting or storing the imagery. A key objective when compressing is to reduce the number of bits necessary to describe the image while maintaining as much of the original "quality" of the image as possible. The definition of quality and the level of quality that is deemed acceptable are application- and user-dependent. Figure 2-1 shows a number of key parameters that figure prominently in determining the amount of entropy or information that is contained within a captured image. Images with more entropy are more complex, and therefore harder to compress effectively. Entropy can be increased through complex scenery, such as grass, shimmering water, shrubs, trees, etc. It can also be increased through unwanted parameters, such as noise. Figure 2-1 provides a look at the prominent "control" variables that effectively act as "tuning" knobs for configuring and managing a streaming or compression session. While not all of these variables can be controlled, it is important to understand how each contributes to the entropy contained within a resulting image, and therefore the effect that each has on the compressibility of the image. In cases where a system is already deployed and being used, there may be limited access to these tuning knobs, and therefore very little flexibility in optimizing for a certain situation. In cases where a system is being developed or initially configured, it is important to consider the flexibility that can be obtained if these tuning knobs are made available within the system.

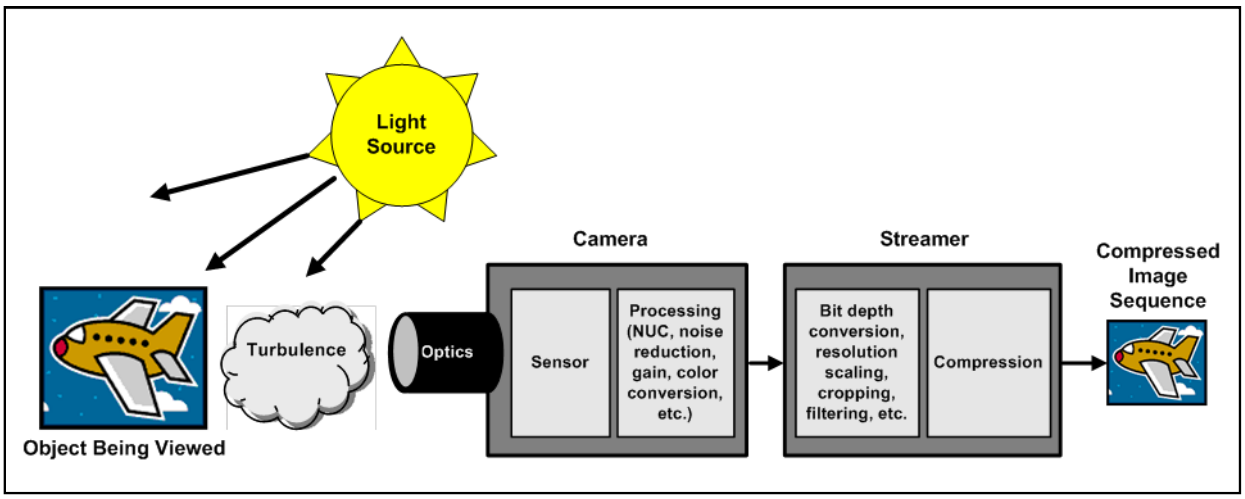

Figure 2-1. Variables that contribute to the content and resultant quality of a compressed image sequence.

The following sections provide a discussion on each of the items listed in Figure 2-1. The information can be used to help in optimizing and configuring a system to achieve expected results. It is important to have a thorough understanding of this system process, as results can vary greatly with subtle changes in a scene or sub-system configuration.

2.1 Light Source

Images are formed via the capture of light. The source of light therefore has a direct bearing on the structure of an image that is created. Two key parameters of light that affect the eventual makeup of an image include the level or amount of light and the spectrum of the light that is absorbed in a sensor. The amount of light affects exposure time settings, and therefore motion blur, contrast, and signal-to-noise ratio. While motion blur tends to decrease the quality of an image, it also tends to make intraframe compression more efficient. An increase in contrast or signal-to-noise increases the quality of an image, but at the same time makes compression less efficient.

The spectrum of the incoming light defines whether an image is captured in the visible or one of the infrared bands. Figure 2-2 shows the most frequently-used imaging bands and how they relate to the transmission characteristics of the atmosphere. Visible band imagery is typically captured in color, although certain range visible cameras capture images in monochrome. In addition, infrared imagery is almost always captured as monochrome imagery. As H.264 is primarily geared toward broadcast and consumer applications, the majority of profiles that are supported are focused on the compression of color imagery.

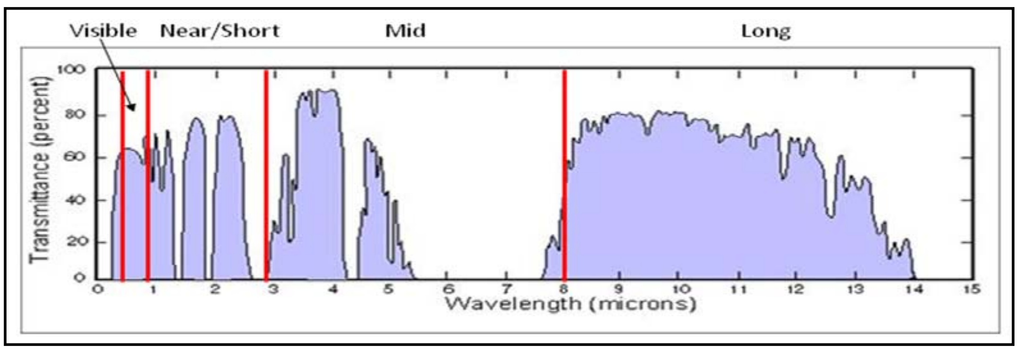

Figure 2-2. Atmospheric transmission and the key T&E imaging bands.

As a result, when leveraging the baseline codec, for instance, monochrome imagery must be formatted into a color format before compression occurs. For infrared imagery, this conversion to color often involves the process of pseudocoloring, which applies a color palette to the monochrome imagery. If the desire is to retain the monochrome appearance of the image, it is still required to be formatted into a color space supported by the codec in use. The choice of color palette and method of color conversion will have an impact on the efficiency and end result from the encoding process. Figure 2-3 shows an infrared image that has been pseudocolored with two different palettes.

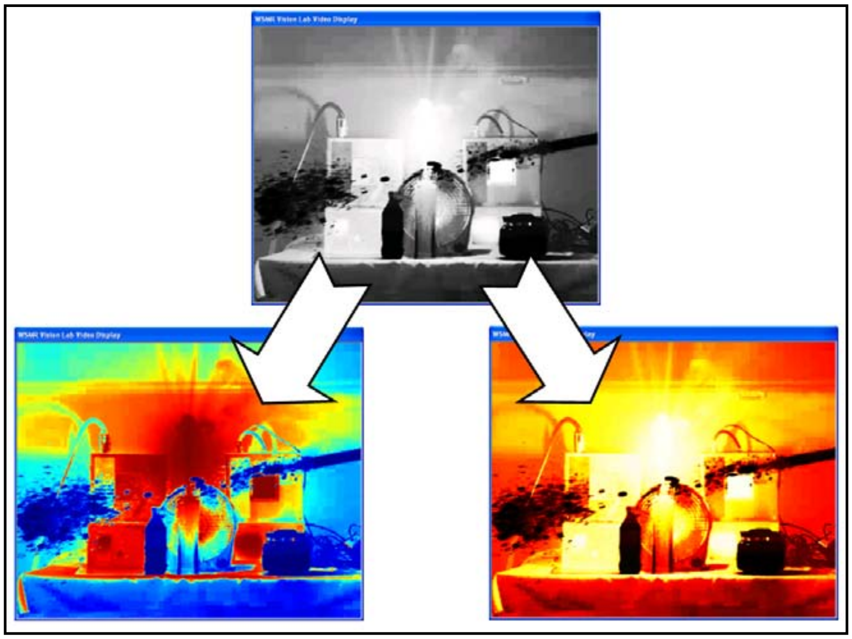

Figure 2-3. Pseudocoloring of monochrome imagery.

2.2 Object/Scene Being Viewed

The complexity of a scene or object being viewed has a direct bearing on the frequency content contained within an image. Figure 2-4 shows two mid-wave infrared images. The image on the left contains a significant amount of plain sky, which is relatively easy to compress, assuming a clean non-uniformity correction. The image on the right contains a number of people with a significant amount of detail across the entire image. As images with higher frequency content are more difficult to compress, it bears out that scenes with grass, bushes, waves, trees, etc. may require a larger number of bits to compress while maintaining a desired level of quality. As one frequently has little control over the object or scene being imaged, the key takeaway is to ensure enough bits are being applied to meet quality expectations, or conversely understand the compromise in quality that may be experienced given a fixed number of bits that can be applied to the compression. The sections on optics and image processing below contain some recommendations on how tradeoffs can be manipulated to one's benefit when a constrained bandwidth situation is encountered.

Figure 2-4. Comparison of a low-complexity scene with a high-complexity scene.

2.3 Atmospheric Turbulence

Practical T&E imaging typically involves viewing an object or scene through the atmosphere at some reasonable distance. Atmospheric turbulence is a phenomenon whereby varying temperatures and density of air pockets along the path of viewing cause light traveling through the atmosphere to bend. As a result, a certain amount of blurring is instilled in the final images that are captured. This process is depicted in Figure 2-5. This blurring due to the turbulence reduces the Modulation Transfer Function (MTF) of the image, resulting in lower entropy. As a result of the lower entropy, the images contain less information and are generally easier to compress. This process would be similar to running a low-pass filter on an image before compressing it.

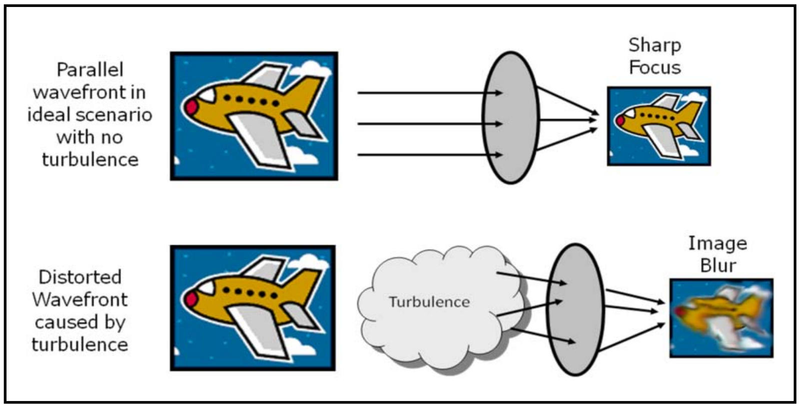

Figure 2-5. Atmospheric turbulence reduces the resolution in captured T&E imagery.

Atmospheric turbulence is a stochastic process and is always changing. It varies with temperature, humidity, barometric pressure, cloud cover, and other factors. Turbulence scales with wavelength, to the 6/5th power. As a result, turbulence in the visible is over ten times worse than would be observed in the mid-wave infrared; therefore, it can be a dominating factor in the visible, while only a minor issue in the mid- or long-wave bands. Figure 2-6 shows a series of turbulence curves for a typical long-range focal length. The 12 cm curve is considered fairly weak turbulence and the 3 cm curve would result in very obvious distortion in the visible band. The straight diagonal curve in the plot represents the diffraction curve for an f/8 optic. As can be seen, the turbulence generally dominates in the visible, while diffraction tends to dominate in the mid-wave infrared.

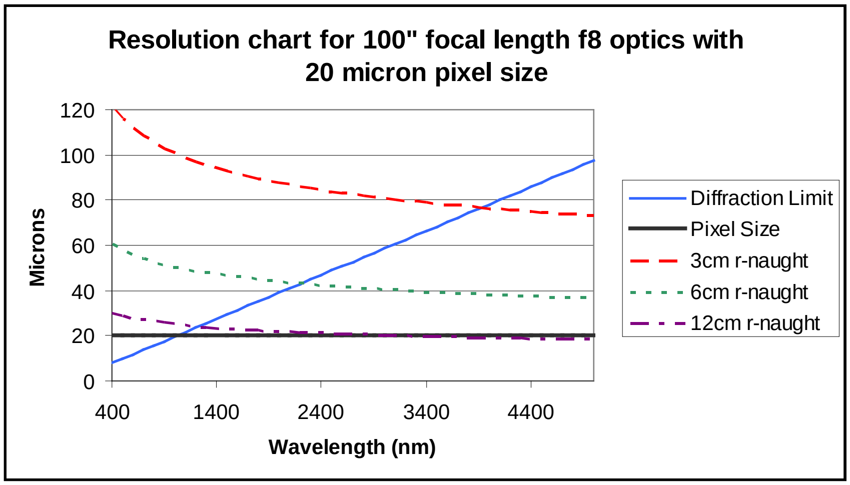

Figure 2-6. Comparison of the effect of atmospheric turbulence and optical diffraction on resolution.

As one does not have the luxury of adjusting the turbulence levels while imaging, it is again important to understand how the process affects the results that are achieved when performing the compression. With respect to compression, atmospheric turbulence will generally reduce the MTF of the image, resulting in images that should compress more efficiently on an intraframe basis. When generating motion-vector based frames, such as H.264 predictive (P) and bi-directional (B) frames, turbulence-induced motion in the image may cause this portion of the compression scheme to become less efficient than it would be without the turbulence.

2.4 Optics and Image Sensor

Both the optics and image sensor of an imaging system have a direct impact on image quality and compression efficiency. For an ideal optic, diffraction determines the finest detail that can be passed through the lens. For non-ideal optics, other types of distortion, such as mis-focus, coma, astigmatism, etc. will degrade the MTF of the image. Degradations in the MTF will tend to make compression more efficient as less information is passed with a lower MTF. Knowledge of this can be useful in certain situations, such as when a camera has poor signal-to-noise or perhaps a bad non-uniformity correction. In these cases, a slight mis-focus or closing of an iris can significantly benefit the encoding process, as the mis-focus or stopped-down iris can drive the lens MTF to a very low number in the area where pixel-to-pixel effects, such as high-frequency noise, are causing difficulty.

As image sensors tend to be very-well-controlled in terms of their geometry, the main non-ideal and detrimental effects include mismatch on pixel gain and/or offset, such as those just mentioned. The quality of non-uniformity correction and color application, such as for a Bayer sensor, can have a strong impact on encoding efficiency and quality. For a given diffraction spot size, smaller pixels will tend to result in an image that is easier to compress. This is due to the fact that the pixel is sampling the incoming wavefront at a lower point on the MTF curve, which means the pixel-to-pixel contrast will generally be lower. The relationship of the pixel size to the optical diffraction spot size typically varies with wavelength band. Table 2-1 shows the

relationship between an ideal diffraction spot size and pixel sizes that are common with many range camera systems.

TABLE 2-1. COMPARISON OF DIFFRACTION SPOT TO PIXEL SIZE			
Waveband	Airy Disc Radius (f/4)	Range of Pixel Size	Ratio of Spot to Pixel
Visible	2.5 µm	~ 5 to 22 µm	0.11 to 0.5
Short-wave Infrared	6.3 µm	~ 15 to 25 µm	0.25 to 0.42
Mid-wave Infrared	20 µm	~ 15 to 25 µm	0.8 to 1.33
Long-wave Infrared	45 µm	~ 25 to 30 µm	1.5 to 1.8

As the size of the pixel increases relative to the size of the optical spot, the pixel samples the optical wavefront at a higher level of the MTF. This creates a higher pixel-to-pixel delta, or edge. With this higher edge strength comes higher image entropy, which becomes more difficult to compress. To show this effect, a series of images of the same scene were captured in the short-wave, mid-wave, and long-wave bands. They are shown in Figure 2-7.

Figure 2-7. Common imagery captured in the short-, mid-, and long-wave bands.

These images were compressed using H.264 in IFrame and 16-frame Intraframe-Bidirectional-Bidirectional-Predictive (IBBP) modes and with JPEG 2000. The relative quality of the compression in each waveband is then shown for each encoding configuration in Figure 2-8, Figure 2-9, and Figure 2-10. For the image quality plots, the encoding bit rate is shown along the X-axis and the peak signal-to-noise ratio (PSNR) (image quality) is shown on the Y-axis. Higher levels of PSNR translate to higher levels of quality for the encoding results. For each of the three compression schemes, the quality of the results follows the pixel-to-optical spot size ratios found in Table 2-1. This fact can be put to good use when configuring multiple channels of video with multi-wavelength content, such as allocating more bandwidth to visible and short-wave channels, and less to mid- and long-wave video channels (for similar array sizes, bit depths, etc.).

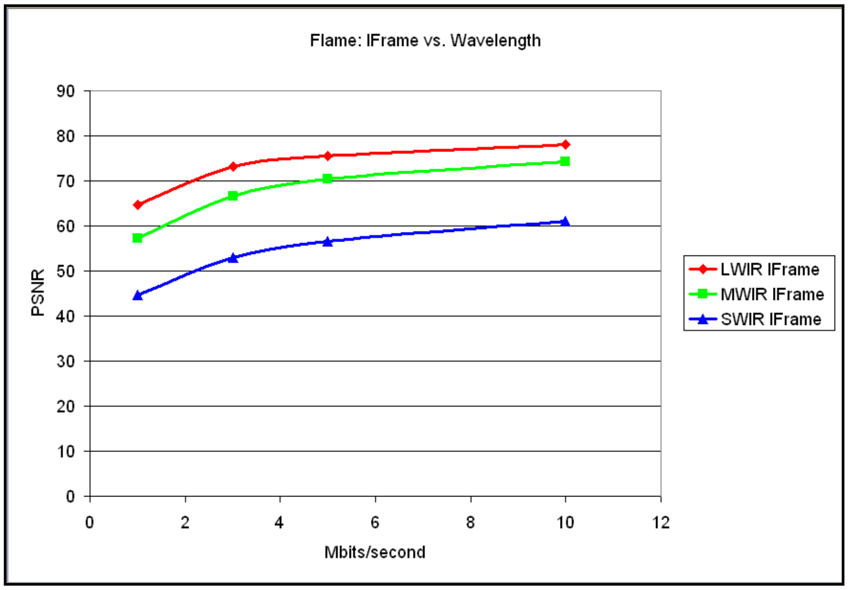

Figure 2-8. Encoding results of short-wave (blue line), mid-wave (green line), and long-wave (red line) imagery using H.264 IFrame.

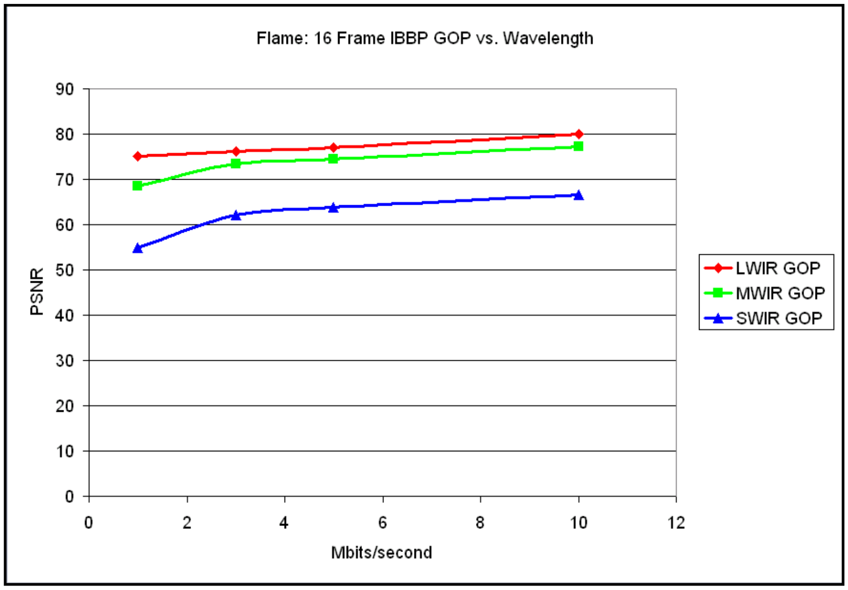

Figure 2-9. Encoding results of short-wave (blue line), mid-wave (green line), and long-wave (red line) imagery using H.264 16-frame IBBP.

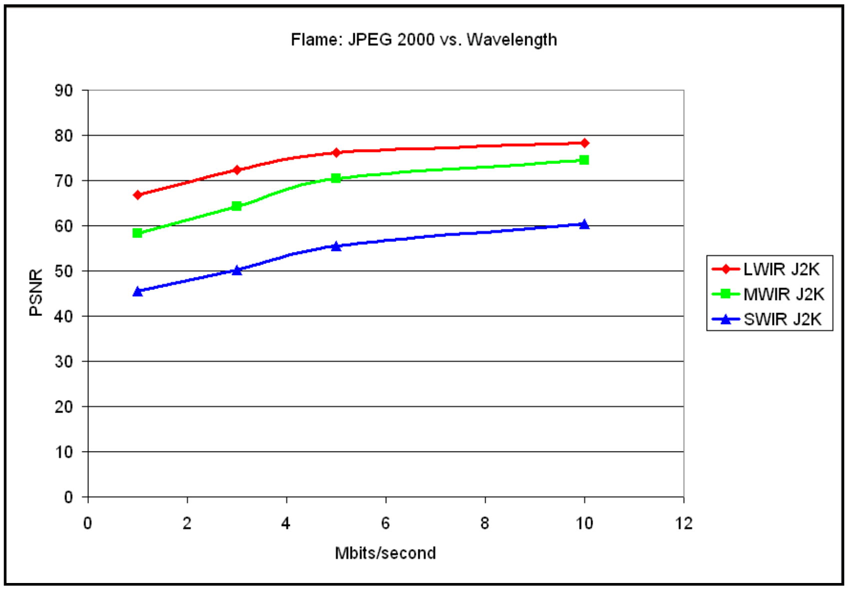

Figure 2-10. Encoding results of short-wave (blue line), mid-wave (green line), and long-wave (red line) imagery using JPEG 2000.

2.5 Image Processing or Manipulation

Cameras and encoders frequently have features that allow for processing, manipulation, or enhancement of the imagery before it is compressed. These operations may include non-uniformity correction, gain, noise reduction, color conversion (generation of 4:2:2, etc.), bit depth conversion, equalization, deconvolution, image resolution scaling, cropping, and filtering. These all have a direct bearing on the encoding process and results. The Engineering Guideline MISB EG0904[2] includes guidelines for how the image encoding process can be configured to account for these types of issues. Table 2-2 shows a brief summary of simple mechanisms for manipulating an image sequence to make it more amenable for compression at a certain desired bandwidth.

TABLE 2-2. MECHANISMS FOR REDUCING BIT RATE OTHER THAN TUNING THE COMPRESSION CODEC DIRECTLY

Method of Bit Rate Reduction	Quality Effect on Result
Binning	Lowers pixel resolution by a factor of four for each level of binning. Causes an increase in neighboring pixel contrast, resulting improvement in bit rate reduction less than a factor of four.
Windowing (Cropping)	Narrows the overall field of view while keeping the same pixel field of view. Acceptable if desired scene content remains within field of view.
Low-pass Filtering	Lowers high-frequency content and softens edges. Can be effective in instances with poor non-uniformity correction.

[2] MISB (Motion Imagery Standards Board). H.264 Bandwidth/Latency/Quality Tradeoffs. EG 0904. Accessed May 2009. Available at http://www.gwg.nga.mil/misb/engpubs.html.

Conversion to 8-bit Pixels From Between 10-bit and 14-bit Pixels	Cuts the byte count for the image in half. Can be very effective for infrared images with low scene dynamic range. For scenes with high dynamic range, scaling conversion may result in loss of desired detail.
Frame Rate Reduction	Throws away temporal information. Effective for scenes with content that has low temporal bandwidth. For scenes with high temporal scene motion, lower frame rates can cause some challenges and inefficiencies for codecs that are configured with a group of pictures (GOP) structure.

For the case of image binning, it should be noted that there is not a one-to-one reduction in encoding bandwidth when imagery is binned. For instance, if an image is binned 2X in both the X and Y direction, resulting in a 4X reduction in the number of pixels, the reduction in bandwidth on the compression side (with similar quality settings) will be less than 4X. The reason for this can be seen in Figure 2-11. When pixels are binned, the resulting "super" pixel samples the incoming image wavefront at a higher point on the MTF curve. This results in higher pixel-to-pixel edge content, making the imagery more difficult to compress.

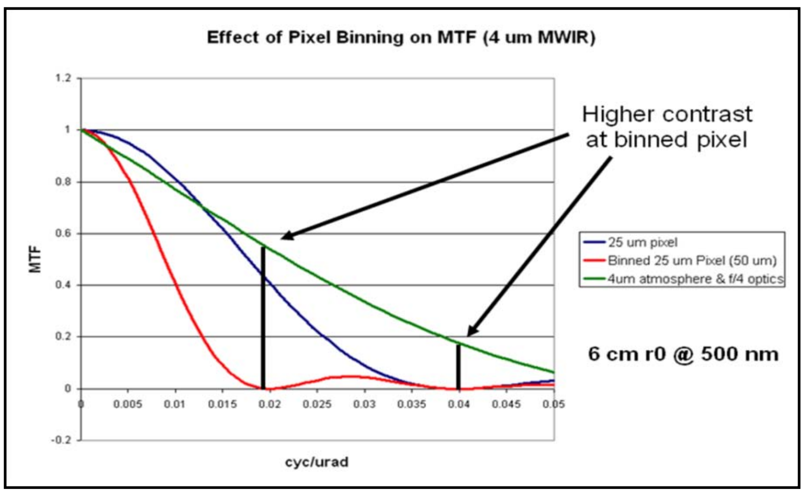

Figure 2-11. The effect of binning and how it affects the sampling relationship with the optical MTF.

2.6 Bit Depth

Commercial and broadcast systems typically operate on 8-bit imagery. Certain higher-end broadcast systems are capable of operating at 10 bits. Range systems, especially those in the infrared, frequently operate at bit depths of up to 14 bits. The majority of commercial encoders operate at 8 bits and the most common profiles for H.264, Baseline, and Hi only support 8-bit imagery. Higher-level profiles, such as the H.264 Fidelity Range Extensions (FRExt) and JPEG 2000, do support range imagery with 14-bit content. The main issue with higher bit depth compression is the requirement for specialized systems that support these infrequently-used profiles.

2.7 Color Format

To benefit the user, a short description of color formats is provided. The formatting and sub-sampling of color information is a standard process with compression systems and is leveraged due to the human eye's inability to distinguish the small changes in color information that result from the sub-sampling. When imagery is captured on a camera sensor, it is captured in a red green blue (RGB) format. In an RGB image, each pixel contains separate values for red, green, and blue. For an 8-bit system, each pixel therefore contains 24 bits of information. Figure 2-12 shows an example of an RGB image.

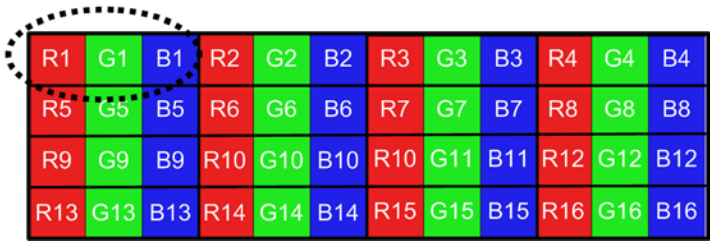

Figure 2-12. A 4x4 block of RGB pixels in an RGB-formatted image.

To perform compression, the RGB image is transformed into YUV or YCbCr color space. The transformation is accomplished via matrix multiplication and varies slightly depending on the source standard for the video imagery. Figure 2-13 displays the conversion matrices for the RGB to YUV and RGB to YCbCR conversions.

$$\begin{bmatrix} Y \\ U \\ V \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ -0.147 & -0.289 & 0.436 \\ 0.615 & -0.515 & -0.100 \end{bmatrix} * \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

RGB to YUV color conversion for analog NTSC/PAL

$$\begin{bmatrix} Y \\ Cb \\ Cr \end{bmatrix} = \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix} + \begin{bmatrix} 0.257 & 0.504 & 0.098 \\ -0.148 & -0.291 & 0.439 \\ 0.439 & -0.368 & -0.071 \end{bmatrix} * \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

RGB to YCbCr color conversion for digital SDTV

$$\begin{bmatrix} Y \\ Cb \\ Cr \end{bmatrix} = \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix} + \begin{bmatrix} 0.183 & 0.614 & 0.062 \\ -0.101 & -0.339 & 0.439 \\ 0.439 & -0.399 & -0.040 \end{bmatrix} * \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

RGB to YCbCr color conversion for digital HDTV

Figure 2-13. Color space conversion.

When an RGB image is transformed into the YCbCr (or YUV) color space, there is initially a one-to-one relationship in terms of the number of parameters used to describe the color space. The advantage of the YCbCr color space is that the eye is not as sensitive to the CbCr terms as it is the Y term. The CbCr terms therefore can be sub-sampled without causing an apparent drop in quality for the image. A source image with all components and values present is referred to as 4:4:4. In this case, for a two-by-two block of pixels, each component has four

values that are represented. In the case of 4:2:2 video, the Cb and Cr terms are sub-sampled in the horizontal direction, which results in four Y terms and two each of Cb and Cr. For 4:2:2 video, 33 percent of the information has been thrown away. Figure 2-14 shows the arrangement of pixels in a 4:2:2 configuration.

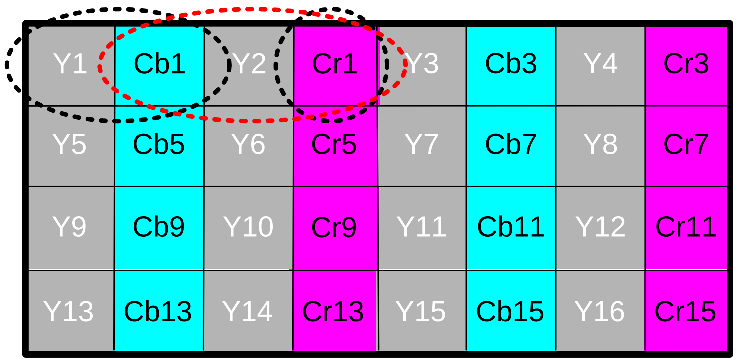

Figure 2-14. A 4x4 frame of 4:2:2 packed YCbCr pixels.

When constructing an image from the terms shown in Figure 2-14, the first pixel is generated by combining Y1, Cb1, and Cr1. The second pixel is generated by combining Y2, Cb1, and Cr1, and so on as shown in Figure 2-15.

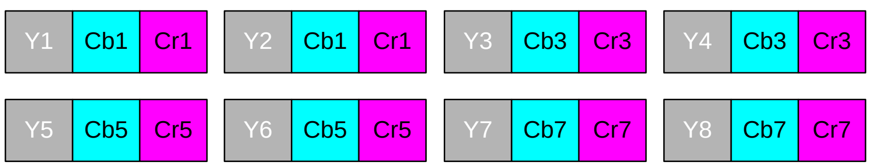

Figure 2-15. Sharing of color information among horizontally-neighboring pixels with 4:2:2 encoding.

In the case of 4:2:0 encoding, which is required in the H.264 baseline profile, the process of “throwing away” color difference information is continued into the vertical direction. The 4:2:0 color information is arranged in a planar configuration in the example shown in Figure 2-16. In this mode, all of the intensity (Y) values are grouped together and the two separate color difference channels are grouped together. For a 2x2 block of pixels, there are four intensity values and one each of Cb and Cr. In this case, half of the original source information has been thrown away in a lossy but efficient manner.

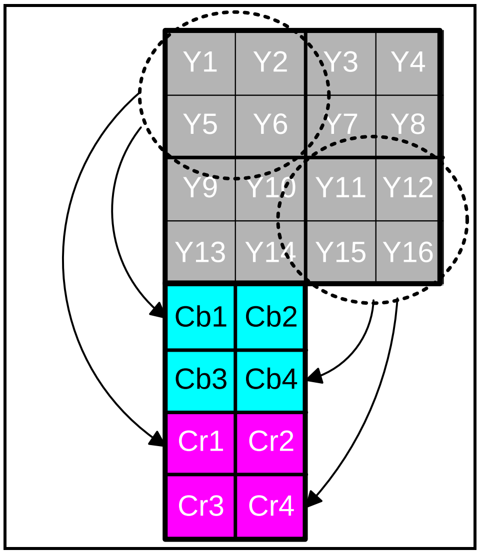

Figure 2-16. 4x4 frame of 4:2:0 planar arranged YCbCr pixels.

When constructing an image from the terms shown in Figure 2-16, the pixels are arranged in the first 2x2 block by combining the Y1, Y2, Y5, and Y6 intensity values with the Cb1 and Cr1 color difference values. The second 2x2 block is generated by combining Y3, Y4, Y7, and Y8 with Cb2 and Cr2, and so on as shown in Figure 2-17.

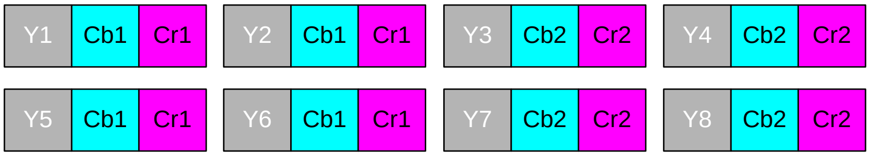

Figure 2-17. Sharing of color information among horizontal and vertical neighboring pixels with 4:2:0 encoding.

CHAPTER 3

COMPRESSION

This section provides information on the H.264 and JPEG 2000 encoding systems and recommendations on profiles and configuration for range use. As compression standards, software, and system capabilities are always being updated and improved, the reader is encourage to seek additional information about current capabilities and the availabilities of commercial and DoD solutions, as well as recommendations and standards that are available through the MISB.

3.1 H.264

The H.264 standard includes a number of different profiles with varying capabilities.[3] Each is targeted toward specific end-user applications. The following are key profiles that are of interest for range applications.

Constrained Baseline Profiles and Baseline Profiles (CBP & BP): These profiles are related and are the most common in use. They are found in many consumer and other low-cost applications, such as videoconferencing and mobile devices. The baseline profile has features that are designed to increase robustness when operating in an environment where data loss might be expected, such as with transmission through a radio frequency (RF) system.

Extended Profile (XP): This profile is leveraged in streaming video applications. It supports high-compression ratios and functions well in applications that experience temporary data loss. The ability to generate and code B frames is added within this profile. While B frames provide for significant bandwidth reduction, they will find limited use in range applications as they result in significant latency when used in streaming applications.

High Profile (HiP): This is the primary profile for high-definition television and for broadcast and disc storage applications. Consumer Blu-ray discs currently store content using this profile. HiP adds support for native coding of 4:0:0 monochrome video.

High 10 Profile (Hi10P): This profile adds 10-bit pixel sampling to HiP. It is mainly used in professional applications and can provide a performance boost in range applications where additional bit depth is desired.

High 4:2:2 Profile (Hi422P): This profile adds support for 4:2:2 chroma sub-sampling to Hi10P. It also supports 10-bit pixel sampling and is targeted at professional interlaced video applications. As range systems, and DoD video systems in general, are targeted more toward progressive scan systems, this profile should see limited use in range applications.

[3] Wikipedia. "H.264/MPEG-4 AVC." Last modified May 23, 2012. http://en.wikipedia.org/wiki/H.264

High 4:4:4 Predictive Profile (Hi444PP): This profile is also known as FRExt and adds support for 4:4:4 chroma sampling, pixel bit depths up to 14 bits, efficient lossless region coding, and the coding of each picture as three separate color planes. This profile holds significant promise for range applications involving higher bit depth camera systems, such as 14-bit infrared. As this profile is considered a niche capability, the majority of commercial systems do not support this profile.

In addition to these profiles, there are constrained versions of the High 10 (**High 10 Intra)**, High 4:2:2 (**High 4:2:2 Intra)**, and High 4:4:4 (**High 4:4:4** and **Context-adaptive Variable-length Coding (CAVLC) 4:4:4)** profiles that are intended for high-end, professional applications. These profiles are of specific interest to the T&E community and are applicable when image quality and the faithful reproduction of the source material are of paramount importance. Intraframe (I) based profiles compress each frame independently and do not produce the motion vector artifacts and errors that are generated by P and B frame processing.

In addition to these profiles, there are additional profiles that have been defined for scalable video coding applications and for multi-view applications, such as stereo 3D. As these currently have limited use in the T&E environment, they will not be discussed here. Table 3-1 provides a summary of some of the key differences between the various H.264 profiles. While Baseline and High are common in consumer applications, the High 4:4:4 profile is the only current profile that supports the higher-end image format types that are commonly generated by T&E high-resolution, high-speed, high-bit depth, and infrared camera systems. To leverage H.264 tools with T&E imagery in its native format requires identifying appropriate niche market tools that support the High 4:4:4 profile. The alternative requires proper scaling and formatting of the imagery so that it conforms to the color and bit depth requirements of one of the more common commercial or broadcast formats.

TABLE 3-1. COLOR FORMAT AND BIT DEPTH DIFFERENCES BETWEEN H.264 PROFILES

	Baseline	**Extended**	**Main**	**High**	**High 10**	**High 4:2:2**	**High 4:4:4 Predictive**
Color Format							
4:0:0 Monochrome Format	No	No	No	Yes	Yes	Yes	Yes
4:2:0 Chroma Format	Yes	Yes	Yes	Yes	Yes	Yes	Yes
4:2:2 Chroma Format	No	No	No	No	No	Yes	Yes
4:4:4 Chroma Format	No	No	No	No	No	No	Yes
Bit depth							
Supported bit depths	8	8	8	8	8 to 10	8 to 10	8 to 14

3.2 JPEG 2000

The JPEG 2000 image coding system[4] consists of 14 separate sub-parts. These parts consist of items such as the core coding system (Part 1), file formats, security, metadata, and others. JPEG 2000 supports both monochrome and high-bit depth imagery, which makes it useful for many range applications. It supports both lossy and lossless compression. It utilizes the discrete wavelet transform (DWT) and produces less harsh artifacts at high compression ratios than encoding systems that leverage the discrete cosine transform (DCT), such as MPEG2 and H.264. JPEG 2000 also performs well on very-high-resolution imagery, such as at high-definition (HD) and above. JPEG 2000 should be considered strongly for use in range archival systems and for systems that can benefit from lossless or visually lossless compression. In systems that download high-resolution imagery over a range network, the lossless (2:1 to 3:1) or visually lossless (up to 10:1) settings can significantly shorten download times. Archival, presentation, and Time Space and Position Information (TSPI) processing are all common motion imagery tasks that can benefit from this capability.

[4] ITU-T. Information Technology – JPEG 2000 image coding system: Core coding System. ITU-T Rec. T.800 (08/02) | ISO IEC 15444-1:2004. n.d. May be superseded by update. Available to ITU-T members and other subscribers at http://www.itu.int/rec/T-REC-T.800-200208-I/en.

CHAPTER 4

INFRARED COMPRESSION

Of specific interest to the range community, and generally not to the commercial or broadcast community as a whole, is that of monochrome and high-bit depth compression. Infrared systems currently in use at the ranges generate raw imagery that is 14-bit monochrome. The MISB has developed a standard, STD0404[5], that details requirements for the application of infrared compression within the DoD. Table 4-1 summarizes the various levels for implementing infrared compression within the DoD, and therefore range community.

TABLE 4-1. MISP INFRARED COMPRESSION LEVELS

IR Compression Options	Compliance Description	Input Color Format	Codec Implementation
Level 1	Minimally Compliant	8-bit 4:2:0	Scaled and converted to 8-bit 4:2:0 color and compressed with MPEG2
Level 2	Less Compliant	8-bit 4:2:0	Scaled and converted to 8-bit 4:2:0 color and compressed with H.264
Level 3	More Compliant	10-bit 4:2:2	Scaled and converted to 10-bit 4:2:2 color and compressed with H.264
Level 4	Fully Compliant	14-bit 4:0:0	JPEG 2000 or H.264 FRExt Profile IDC Level 244, High 4:4:4

4.1 Comparing Infrared Compression with H.264 and JPEG 2000

As a general rule of thumb, H.264 tends to provide strong performance when encoding standard broadcast format motion imagery. JPEG 2000 tends to provide strong performance when encoding high-resolution and/or high-bit depth imagery. Figure 4-1 contains the results of H.264 and JPEG 2000 compression on source imagery that consists of 14-bit monochrome mid-wave images of a lab scene that contains a torch, some blackbodies, and a coffee pot and plastic bottle filled with ice water.

[5] MISB (Motion Imagery Standards Board). Compression for Infrared Motion Imagery. STANDARD 0404. n.d. May be superseded by update. Available at http://www.gwg.nga.mil/misb/stdpubs.html.

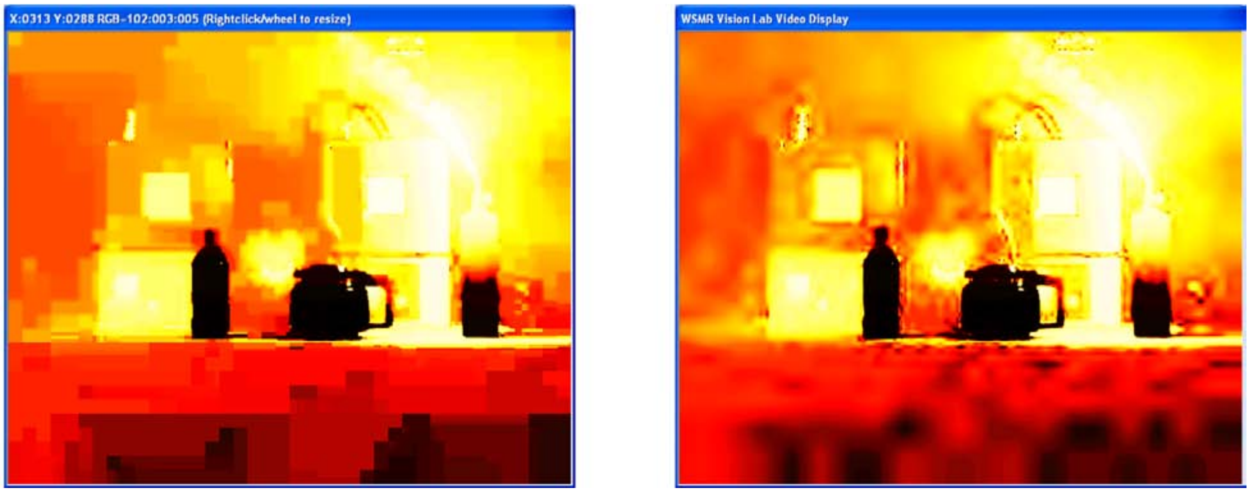

Figure 4-1. Comparison of compression artifacts with highly compressed 14-bit intraframe H.264 and JPEG 2000 infrared imagery.

The source imagery consists of 640x512, 14-bit monochrome 4:0:0 images. In Figure 4-1, they have been compressed to approximately 138 to 1 in intraframe mode. This compression is higher than would normally be used in practice. It is used here to highlight the structure of the breakdown within each algorithm. The results have been pseudocolored to enhance the appearance of the artifacts. With the H.264 result, the main artifacting is the blockiness due to the DCT. With the JPEG 2000 result, the artifacting is more of a filter ringing behavior, which results from the use of the DWT.

Figure 4-2 through Figure 4-5 show the gradual decline in performance starting with reasonable compression at 13 to 1 in Figure 4-2 and increasing to 137 to 1 in Figure 4-5.

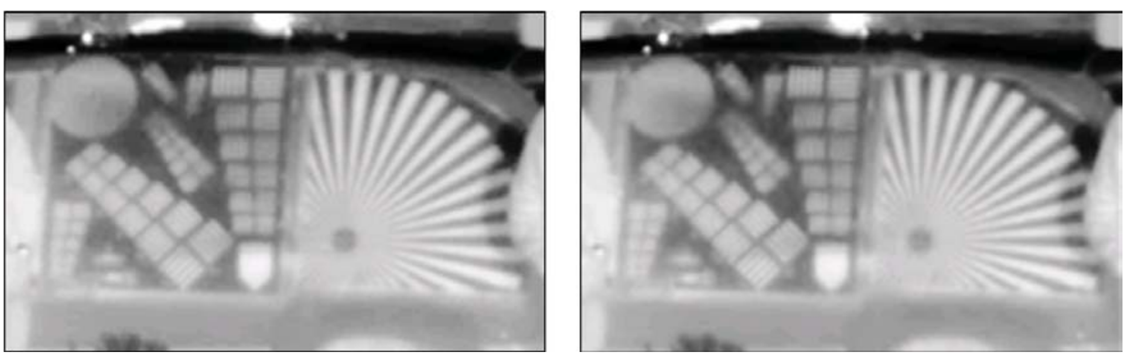

Figure 4-2. H.264 vs. JPEG 2000 at 13 to 1 (10 Mbps stream equivalent).

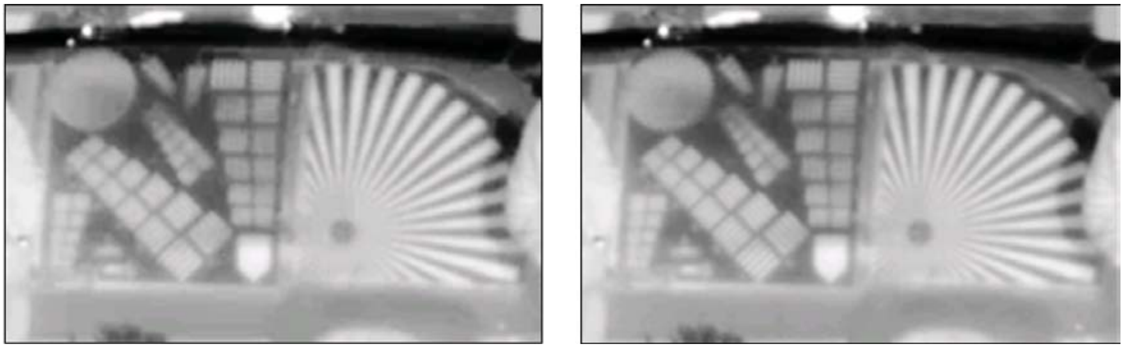

Figure 4-3. H.264 vs. JPEG 2000 at 26 to 1 (5 Mbps stream equivalent).

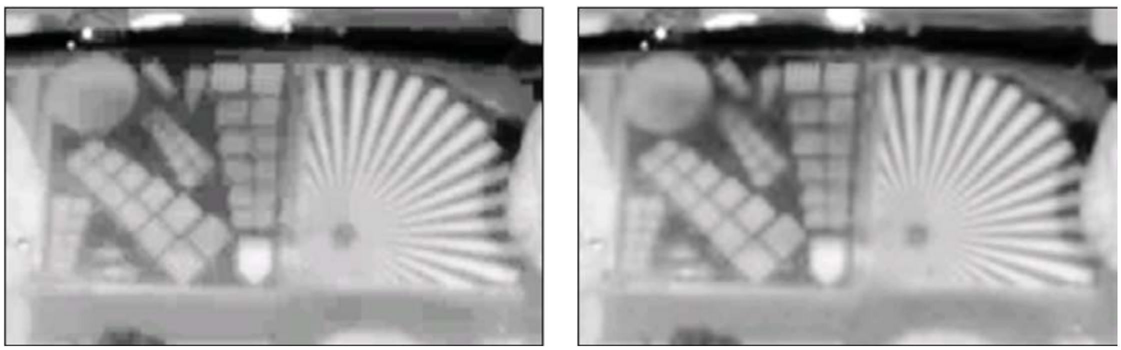

Figure 4-4. H.264 vs. JPEG 2000 at 46 to 1 (3 Mbps stream equivalent).

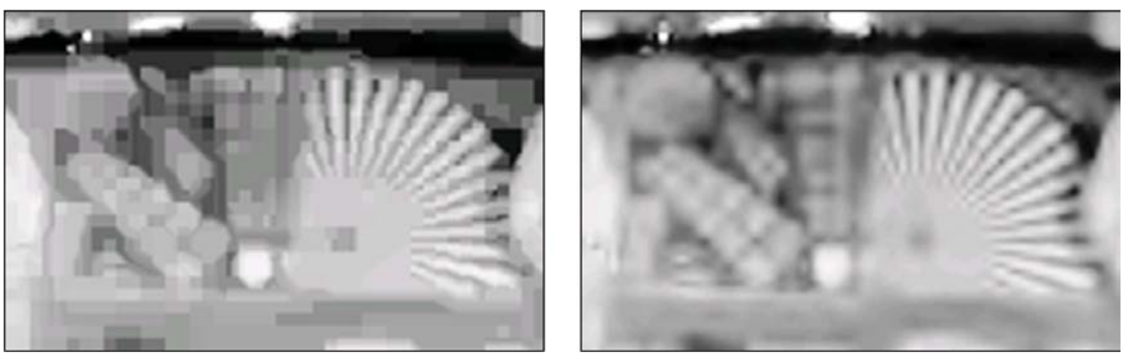

Figure 4-5. H.264 vs. JPEG 2000 at 137 to 1 (1 Mbps stream equivalent).

While the results in Figure 4-5 indicate a distinct advantage for JPEG 2000, it needs to be noted that the JPEG 2000 codec[6] used in this comparison is much more mature than the H.264 FRExt codec[7] that was used. The H.264 results are expected to improve as this capability matures.

[6] Kakadu Software website; http://www.kakadusoftware.com

[7] JM Reference Software website; http://iphome.hhi.de/suehring/tml/

CHAPTER 5

RECOMMENDATIONS

Image compression and streaming is a complex technology that requires a sufficient understanding to properly implement in range applications. Understanding the requirements of an application and applying the correct codec and configuration is critical for success. Range applications vary in their compression requirements.

The requirements for compression are driven by the make-up of the source imagery and the need for bandwidth reduction and desired minimum image quality. Bandwidth reduction is driven by requirements such as file format size and link or network transmission constraints. Image quality requirements are driven by the use of the video. Table 5-1 shows common range compression and streaming applications and recommendations for configuration of those systems.

TABLE 5-1. CONFIGURATION RECOMMENDATIONS FOR RANGE COMPRESSION SYSTEMS

General Application	Codec Guidelines	Compression Ratios	Comments
Real-time low latency critical streaming	- Intraframe (if bandwidth allows) - IPP (4 to 6 P frames) - Avoid the use of B frames due to latency and transient event issues	20:1 to 40:1	-Man-in-the-loop control - Safety - Live monitoring - Intraframe supports optimum capture of fast transient events - IPP GOP optimizes bandwidth while keeping latency low
Real-time non-critical streaming in a bandwidth-constrained environment	- Intraframe with binning, windowing, or reduced frame rate - IPP (4 to 6 P frames) with binning, windowing, or reduced frame rate	40:1 to 90:1	- General instrumentation observation - VIP viewing
Non-real-time, highly constrained bandwidth distribution	- H.264 with IBBP GOP	60:1 to 120:1	- Non-real-time playback over distributed networks
Post test analysis	H.264 IFrame or JPEG 2000	Lossless to 10:1	- Retention of TSPI and/or radiometric calculation accuracy - Desire single-frame access for review

Post-test, quick look presentation and review	- Intraframe if bandwidth allows - IPP (4 to 6 P frames)	20:1 to 40:1	- Desire fine-grain frame access for freeze frame and review
Archival	H.264 IFrame or JPEG 2000	Lossless to 10:1	- Retention of TSPI and/or radiometric calculation accuracy - Desire single-frame access for review

5.1 T&E Compression & Bit Rate Matrix

Table 5-2 through Table 5-7 represent reasonable compression and bit rate values for various broadcast format video streams. It should be noted that as the implementation of compression codecs evolve, they become more efficient over time. As a result, the values in these tables will likely evolve and improve over time. For those tasked with configuring range systems, it is recommended to monitor the capabilities of fielded and upgraded systems to ensure they are configured in an optimum manner.

TABLE 5-2. COMPRESSION RATES VERSUS TYPICAL QUALITY AND LATENCY

Resolution	Raw	10:1	30:1	90:1	120:1
Sensor Width, Height, Frame Rate	Uncompressed Bandwidth	Visually Lossless, <200 ms latency	High quality, <200 ms latency	Moderate quality, 0.5s – 1.0 s latency	Distributed SA, presentation, > 1.0 s latency

TABLE 5-3. VARIOUS 4:4:4 COLOR MOTION IMAGERY FORMATS AND THEIR ASSOCIATED BIT RATES – FULL COLOR

Resolution	Raw	10:1	30:1	90:1	120:1
640x480@30	211 Mbps	21 Mbps	7 Mbps	2.3 Mbps	1.8 Mbps
1280x720@30	633 Mbps	63 Mbps	21 Mbps	7.0 Mbps	5.3 Mbps
1280x720@60	1266 Mbps	127 Mbps	42 Mbps	14.1 Mbps	10.6 Mbps
1920x1080@30	1424 Mbps	142 Mbps	47 Mbps	15.8 Mbps	11.9 Mbps
1920x1080@60	2848 Mbps	285 Mbps	94 Mbps	31.6 Mbps	23.7 Mbps

TABLE 5-4. VARIOUS 4:2:2 COLOR MOTION IMAGERY FORMATS AND THEIR ASSOCIATED BIT RATES – BROADCAST COLOR

Resolution	Raw	10:1	30:1	90:1	120:1
640x480@30	141 Mbps	14 Mbps	4.7 Mbps	1.6 Mbps	1.2 Mbps
1280x720@30	422 Mbps	42 Mbps	14 Mbps	4.7 Mbps	3.5 Mbps
1280x720@60	844 Mbps	84 Mbps	28 Mbps	9.4 Mbps	7.0 Mbps
1920x1080@30	949 Mbps	95 Mbps	32 Mbps	11 Mbps	7.9 Mbps
1920x1080@60	1898 Mbps	190 Mbps	63 Mbps	21 Mbps	16 Mbps

TABLE 5-5. VARIOUS 4:2:0 COLOR MOTION IMAGERY FORMATS AND THEIR ASSOCIATED BIT RATES – CONSUMER COLOR

Resolution	Raw	10:1	30:1	90:1	120:1
640x480@30	105 Mbps	11 Mbps	3.5 Mbps	1.2 Mbps	879 kbps
1280x720@30	316 Mbps	32 Mbps	11 Mbps	3.5 Mbps	2.6 Mbps
1280x720@60	633 Mbps	63 Mbps	21 Mbps	7.0 Mbps	5.3 Mbps
1920x1080@30	712 Mbps	71 Mbps	24 Mbps	7.9 Mbps	5.9 Mbps
1920x1080@60	1424 Mbps	142 Mbps	47 Mbps	16 Mbps	12 Mbps

TABLE 5-6. VARIOUS 10- TO 16-BIT MONOCHROME MOTION IMAGERY FORMATS AND THEIR ASSOCIATED BIT RATES

Resolution	Raw	10:1	30:1	90:1	120:1
640x480@30	141 Mbps	14 Mbps	4.7 Mbps	1.6 Mbps	1.2 Mbps
1280x720@30	422 Mbps	42 Mbps	14 Mbps	4.7 Mbps	3.5 Mbps
1280x720@60	844 Mbps	84 Mbps	28 Mbps	9.4 Mbps	7 Mbps
1920x1080@30	949 Mbps	95 Mbps	32 Mbps	11 Mbps	7.9 Mbps
1920x1080@60	1898 Mbps	190 Mbps	63 Mbps	21 Mbps	16 Mbps

TABLE 5-7. VARIOUS 8-BIT MONOCHROME MOTION IMAGERY FORMATS AND THEIR ASSOCIATED BIT RATES

Resolution	Raw	10:1	30:1	90:1	120:1
640x480@30	70 Mbps	7 Mbps	2.3 Mbps	781 kbps	586 kbps
1280x720@30	211 Mbps	21 Mbps	7 Mbps	2.3 Mbps	1.8 Mbps
1280x720@60	422 Mbps	42 Mbps	14 Mbps	4.7 Mbps	3.5 Mbps
1920x1080@30	475 Mbps	47 Mbps	16 Mbps	5.5 Mbps	4 Mbps
1920x1080@60	949 Mbps	95 Mbps	32 Mbps	11 Mbps	7.9 Mbps

REFERENCES

The following references are current as of the publishing of this document. Versions are expected to be updated over time and the reader is encouraged to stay current with the latest version of these documents. This list of references includes two articles published by Wikipedia. Because Wikipedia was designed so that anybody with access can add or modify its contents, the contents are vulnerable to being changed, misleading, or incorrect.

ITU-T (International Telecommunications Union Telecommunication Standardization Sector). Advanced video coding for generic audiovisual service | Information Technology – Coding of audio-visual objects Part 10: Advanced Video Coding. ITU-T Rec. H.264 (03/09) | ISO/IEC 14496-10:2009. n.d. Superseded by current version available at http://www.itu.int/rec/T-REC-H.264/en.

———. Information technology – Generic coding of moving pictures and associated audio information: Systems. ITU-T Rec.H.222.0 (05/06) | ISO/IEC 13818-1:2007. n.d. May be superseded by update. Available to ITU-T members and other subscribers at http://www.itu.int/rec/T-REC-H.222.0/en.

———. Information Technology – JPEG 2000 image coding system: Core coding System. ITU-T Rec. T.800 (08/02) | ISO IEC 15444-1:2004. n.d. May be superseded by update. Available to ITU-T members and other subscribers at http://www.itu.int/rec/T-REC-T.800-200208-I/en.

Kakadu Software website; http://www.kakadusoftware.com

JM Reference Software website; http://iphome.hhi.de/suehring/tml/

MISB (Motion Imagery Standards Board). Compression for Infrared Motion Imagery. STANDARD 0404. n.d. May be superseded by update. Available at http://www.gwg.nga.mil/misb/stdpubs.html.

———. Constructing a MISP Compliant File. TRM 0909. Accessed September 2009. Available at http://www.gwg.nga.mil/misb/trm.html.

———. Inserting Time Code and Metadata in High Definition Uncompressed Video. RP 0605. Accessed September 2007. Superseded by Inserting Time Stamps and Metadata in High Definition Uncompressed Video. STANDARD 0605.3. Available at http://www.gwg.nga.mil/misb/stdpubs.html.

———. H.264 Bandwidth/Latency/Quality Tradeoffs. EG 0904. Accessed May 2009. Available at http://www.gwg.nga.mil/misb/engpubs.html.

———. H.264/AVC Coding and Multiplexing. EG 0802. Accessed May 2009. Available at http://www.gwg.nga.mil/misb/engpubs.html.

———. Motion Imagery Standards Profile v6.2. Accessed June 2011. Available at http://www.gwg.nga.mil/misb/misppubs.html.

———. Security Metadata Universal and Local Sets for Digital Motion Imagery. STANDARD 0102.7. Accessed Sept 2009. Available at http://www.gwg.nga.mil/misb/stdpubs.html.

———. Surfing the MISP. TRM 1007. Accessed September 2009. Available at http://www.gwg.nga.mil/misb/trm.html.

———. Time Stamping Compressed Motion Imagery. STANDARD 0604. Accessed October 2008. Superseded by Time Stamping and Transport of Compressed Motion Imagery and Metadata. STANDARD 0604.2. Available at http://www.gwg.nga.mil/misb/stdpubs.html.

———. UAS Datalink Local Metadata Set. STANDARD 0601.3. Accessed May 2009. Available at http://www.gwg.nga.mil/misb/stdpubs.html.

Motion Imagery Standards Board website; www.gwg.nga.mil/misb

Wikipedia. "H.264/MPEG-4 AVC." Last modified May 23, 2012. http://en.wikipedia.org/wiki/H.264

———. Plot of Earth atmosphere transmittance in the infrared region of the electromagnetic spectrum. Uploaded November 30, 2005. http://en.wikipedia.org/wiki/File:Atmospheric.transmittance.IR.jpg